Gabor Feature based Robust Representation and Classification for Face Recognition with Gabor Occlusion Dictionary

Meng Yang, Lei Zhang[1], Simon C. K. Shiu, David Zhang

Dept. of Computing, The Hong Kong Polytechnic University, Hong Kong, China

Abstract: By representing the input testing image as a sparse linear combination of the training samples via l_1-norm minimization, sparse representation based classification (SRC) has shown promising results for face recognition (FR). Particularly, by introducing an identity occlusion dictionary to code the occluded portions of face images, SRC could lead to robust FR results against face occlusion. However, the l_1-norm minimization and the high number of atoms in the identity occlusion dictionary make the SRC scheme computationally very expensive. In this paper, a Gabor feature based robust representation and classification (GRRC) scheme is proposed for robust FR. The use of Gabor features not only increases the discrimination power of face representation, but also allows us to compute a compact Gabor occlusion dictionary which has much less atoms than the identity occlusion dictionary. Furthermore, we show that with Gabor feature transformation, l_2-norm could take place the role of l_1-norm to regularize the coding coefficients, which reduces significantly the computational cost in coding occluded face images. Our extensive experiments on benchmark face databases, which have variations of lighting, expression, pose and occlusion, demonstrated the high effectiveness and efficiency of the proposed GRRC method.

Keywords: Robust face recognition, Gabor occlusion dictionary, sparse representation, collaborative representation

---

[1] Corresponding author. Email: cslzhang@comp.polyu.edu.hk.

1. Introduction

Automatic face recognition (FR) is one of the most visible and challenging research topics in computer vision, machine learning and biometrics [11]. Although facial images have a high dimensionality, they usually lie in a lower dimensional subspaces or sub-manifolds. Therefore, subspace learning and manifold learning methods have been dominantly and successfully used in appearance based FR [1-9]. The classical Eigenface and Fisherface [1-4] algorithms consider only the global scatter of training samples and they fail to reveal the essential data structures nonlinearly embedded in high dimensional space. The manifold learning methods were proposed to overcome this limitation [5-6], and the representative manifold learning methods include locality preserving projection (LPP) [7], local discriminant embedding (LDE) [8], unsupervised discriminant projection (UDP) [9], etc. Besides, in order to better exploit the prior knowledge that face images from a single subject could construct a subspace, nearest subspace (NS) classifiers [19][35][36][37] are developed, which are usually superior to the simple nearest neighbor (NN) classifier.

The success of manifold learning implies that the high dimensional face images can be sparsely represented or coded by the representative samples on the manifold. Recently an interesting work was reported by Wright *et al.* [10], where the sparse representation (SR) technique is employed for FR. In Wright *et al.*'s pioneer work, the training face images are used as the dictionary to code an input testing image as a sparse linear combination of them via l_1-norm minimization. The SR based classification (SRC) of face images is conducted by evaluating which class of training samples could result in the minimal reconstruction error of the input testing image with the associated sparse coding coefficients. To make the l_1-norm sparse coding computationally feasible, in general the dimensionality of the training and testing face images should be reduced, or a set of features could be extracted from the original image for SRC. In the case of FR without occlusion, Wright *et al.* [10] tested different types of features, including Eigenface [2], Randomface [10] and Fisherface [3], and they claimed that SRC is insensitive to feature types when the feature dimension is large enough. In the case of FR with occlusion/corruption, an occlusion dictionary was introduced in SRC to code the occluded/corrupted components [10]. Since the occluded face image can be viewed as a summation of non-occluded face image and the occlusion, with the sparsity constraint the non-occluded face part is expected to be sparsely coded by the training face dictionary only, while the occlusion part is

2

expected to be coded by the occlusion dictionary only. Consequently, the classification can be performed based on the reconstruction residuals using the coding coefficients over the training face dictionary. Such a scheme has shown to be effective in overcoming the problem of face occlusion.

The success of SRC boosts the research of sparsity based FR, and many works have been consequently reported. Gao *et al.* [28] proposed the kernel sparse representation for FR and image classification, while FR with continuous occlusion and misalignment via sparse representation have been presented in [29][33][34]. Elhamifar *et al.* [30] discussed classification using structure sparse representation to exploit the block structure of the dictionary, and Yang *et al.* [31] proposed a robust sparse coding model with a maximum likelihood estimator like fidelity term. Moreover, learning a discriminative dictionary under the sparse representation framework for classification has also attracted much attention. Zhang and Li [40] introduced discrimination information into the algorithm of K-SVD [39] by learning a linear classifier; Jiang *et al.* [41] further enhanced dictionary's discriminative ability by adding label consistent information. Very recently, Yang *et al.* [42] imposed Fisher discrimination criterion on the coding residuals and coefficients, and proposed a Fisher discrimination dictionary learning method.

Although the SRC based FR scheme proposed in [10] is very creative and effective, there are two issues to be further addressed. First, the features of Eigenface, Randomface and Fisherface tested in [10] are all holistic features. Since in practice the number of training samples is often limited, such holistic features cannot effectively handle the variations of illumination, expression, pose and local deformations. The claim made in [10] that feature extraction is not so important to SRC actually holds only for holistic features. Second, the occlusion matrix proposed in [10] is an orthogonal matrix, such as the identify matrix, Fourier bases or Haar wavelet bases, etc. However, the number of atoms required in the orthogonal occlusion matrix is very high. For example, if the dimensionality of features used in SRC is 3000, then a 3000×3000 occlusion matrix is needed. Such a big occlusion matrix makes the sparse coding process very computationally expensive, and even prohibitive. These two issues are not fully solved by the sparsity based FR improvers [28-34][40-42]. For instance, only holistic features are considered in [29-34][40-42], FR with occlusion is ignored in [28][32][33], and no occlusion dictionary is learned in [40-42].

3

In our previous work [38], the Gabor features were adopted for SRC and a Gabor occlusion dictionary was learned under the sparse representation framework. Although the so-called Gabor-based SRC scheme improves much the accuracy and efficiency of original SRC, the l_1-norm sparsity constraint on the coding coefficients still makes it time consuming. Very recently, Zhang *et al.* [32] showed that it is the collaborative representation mechanism (i.e., representing the query image collaboratively by samples from all the classes) but not the l_1-norm sparsity constraint on coding coefficients that makes SRC effective for FR. In light of this finding, in this paper we propose a Gabor-feature based robust representation and classification (GRRC) scheme for FR, which will not only be robust to face occlusion but also have much higher computational efficiency than the previous methods such as Gabor-based SRC.

The Gabor filter was first introduced by David Gabor in 1946 [14], and was later shown as models of simple cell receptive fields [15]. The Gabor filters, which could effectively extract the image local directional features on multiple scales, have been successfully and prevalently used in FR [16][17][18]. Very recently, Zhou *et al.* [47] proposed to combine the perceptual features by Gabor filtering with diffusion distance for FR; Du *et al.* [48] proposed to perform FR with non-uniform multilevel selection of Gabor features instead of the uniform down-sampling of Gabor features; a local Gabor based FR with improved accuracy by the selection of Gabor jets was presented in [49]; and multimodal FR using Gabor feature was presented in [50]. All of these methods lead to state-of-the-art results. The local Gabor features are less sensitive to variations of illumination, expression and poses than the holistic features such as Eigenface and Randomface [10]. In the proposed GRRC, the use of Gabor kernels will not only improve much the FR accuracy, it will also allow us to learn a compact occlusion dictionary to deal with face occlusions. Compared with the occlusion dictionary used in SRC, the number of atoms is significantly reduced (often with a ratio of 40:1 ~ 50:1 in our experiments) in the Gabor occlusion dictionary (GOD) used in GRRC. Particularly, it is found that the coding coefficients over the compact GOD can be regularized by l_2-norm, instead of the l_1-norm adopted in Gabor-based SRC [38]. This significantly reduces the computational cost in coding occluded face images. Our experiments on benchmark face databases clearly validate the performance of the proposed GRRC method.

The rest of the paper is organized as follows. Section 2 briefly reviews SRC, collaborative representation based classification (CRC) and Gabor filters. Section 3 presents the proposed GRRC algorithm. Section 4 conducts experiments. Section 5 gives some discussions and Section 6 concludes the paper.

Table 1 summarizes the abbreviations used throughout the paper.

Table 1: Abbreviations used in this paper.

Abbreviation	Meaning
FR	Face Recognition
GOD	Gabor Occlusion Dictionary
GRR	Gabor-feature based Robust Representation
GRRC	Gabor-feature based Robust Representation based Classification
GRRC_L_p	GRRC with l_p-norm regularization
SRC	Sparse Representation-based Classification
CRC	Collaborative Representation based Classification

2. Related Work

2.1. Sparse representation based classification (SRC)

The sparse representation based classification (SRC) method was presented in [10] for robust face recognition (FR). Denote by $A_i = [s_{i,1}, s_{i,2}, ..., s_{i,n_i}] \in \Re^{m \times n_i}$ the set of training samples of the ith object class, where $s_{i,j}, j=1,2,...,n_i$, is an m-dimensional vector stretched by the jth sample of the ith class. For a test sample $y_0 \in \Re^m$ from the ith class, intuitively, y_0 could be well approximated by the linear combination of the samples within A_i, i.e., $y_0 \approx \sum_{j=1}^{n_i} \alpha_{i,j} s_{i,j} = A_i \alpha_i$, where $\alpha_i = [\alpha_{i,1}, \alpha_{i,2}, ..., \alpha_{i,n_i}]^T \in \Re^{n_i}$ is the coding vector. Suppose we have K object classes, and let $A=[A_1, A_2,..., A_K]$ be the concatenation of the n training samples from all the K classes, where $n=n_1+n_2+...+n_K$, then the linear representation of y_0 can be written in terms of all training samples as $y_0 \approx A\alpha$, where $\alpha=[\alpha_1;...;\alpha_i;...; \alpha_K]=[0,...,0, \alpha_{i,1}, \alpha_{i,2},..., \alpha_{i,n_i}, 0,...,0]^T$.

In SRC without occlusion, first y_0 is sparsely coded on A via l_1-minimization

$$\hat{\alpha} = \arg\min_{\alpha} \left\{ \|y_0 - A\alpha\|_2^2 + \lambda \|\alpha\|_1 \right\} \tag{1}$$

where λ is a scalar constant. Then classification is made by

$$\text{identity}(\mathbf{y}_0) = \arg\min_i \{e_i\} \tag{2}$$

where $e_i = \left\| \mathbf{y}_0 - A_i \delta_i(\hat{\boldsymbol{\alpha}}) \right\|_2$, $\hat{\boldsymbol{\alpha}} = \left[\delta_1(\hat{\boldsymbol{\alpha}}); \cdots; \delta_i(\hat{\boldsymbol{\alpha}}); \cdots; \delta_K(\hat{\boldsymbol{\alpha}}) \right]$, and $\delta_i(\cdot): \Re^n \to \Re^{n_i}$ is the characteristic

function which selects the coefficients associated with the ith class.

In SRC with occlusion or corruption, the test sample $\mathbf{y}$ is rewritten as

$$\mathbf{y} = \mathbf{y}_0 + \mathbf{e}_0 = A\boldsymbol{\alpha} + \mathbf{e}_0 = \begin{bmatrix} A, & A_e \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\alpha}_e \end{bmatrix} = B\boldsymbol{\omega} \tag{3}$$

where $B = \begin{bmatrix} A, A_e \end{bmatrix} \in \Re^{m \times (n + n_e)}$, and the clean face image $\mathbf{y}_0$ and the corruption error $\mathbf{e}_0$ have sparse

representations over the training sample dictionary A and occlusion dictionary $A_e \in \Re^{m \times n_e}$, respectively. In

[10], the corruption dictionary A_e was set as an orthogonal matrix, such as identity matrix, Fourier bases,

Haar wavelet bases, and so on. The sparse coding coefficient $\boldsymbol{\omega}$ could be solved via l_1-minimization like Eq.

(1) and the classification is done via Eq. (3) with $e_i = \left\| \mathbf{y} - A_i \delta_i(\boldsymbol{\alpha}) - A_e \boldsymbol{\alpha}_e \right\|_2$.

2.2. Collaborative representation based classification (CRC)

Though it was claimed in [10] that the l_1-norm sparsity imposed on coding coefficient $\boldsymbol{\alpha}$ in Eq. (1) is the key

for the success of SRC, recently it has been shown in [32] that it is the collaborative representation (CR)

mechanism, but not the l_1-norm sparsity on $\boldsymbol{\alpha}$, that truly makes SRC effective for face classification. Using

l_2-norm to regularize $\boldsymbol{\alpha}$ leads to similar FR results. The robustness to outliers in SRC actually comes from

the use of l_1-norm to model the coding error, i.e., $\|\boldsymbol{\alpha}_e\|_1$. Without considering the robustness to outliers, the

coding model of collaborative representation for classification (CRC) [32] is

$$\hat{\boldsymbol{\alpha}}_1 = \arg\min_{\boldsymbol{\alpha}} \left\{ \left\| \mathbf{y}_0 - A\boldsymbol{\alpha} \right\|_2^2 + \lambda \left\| \boldsymbol{\alpha} \right\|_2^2 \right\} \tag{4}$$

The classification of CRC is performed by checking which class yields the minimal regularized coding error,

which is similar to that of SRC.

It is shown in [32] that CRC has very competing performance with SRC in FR without occlusion but with

much faster speed. However, the standard CRC in Eq. (4) does not aim to deal with FR with occlusion.

Compared to CRC [32] which has no occlusion dictionary, in this paper Gabor occlusion dictionary is learnt

to effectively handle occluded portions in facial images. In addition, the use of Gabor features instead of original image intensity also enhances much the discrimination of face representation.

2.3. Gabor features

The Gabor filters (kernels) with orientation μ and scale v are defined as [17]:

$$\psi_{\mu,v}(z) = \frac{\left\|k_{\mu,v}\right\|^2}{\sigma^2} e^{-\left\|k_{\mu,v}\right\|^2\|z\|^2/2\sigma^2} \left(e^{ik_{\mu,v}z} - e^{-\sigma^2/2}\right) \tag{5}$$

where $z=(x,y)$ denotes the pixel, $\|.\|$ denotes the norm operator, and the wave vector $k_{\mu,v}$ is defined as $k_{\mu,v} = k_v e^{i\phi_\mu}$ with $k_v = k_{max}/f^v$ and $\phi_\mu = \pi\mu/8$. k_{max} is the maximum frequency, and f is the spacing factor between kernels in the frequency domain. In addition, σ determines the ratio of the Gaussian window width to wavelength.

Convolving image *Img* with a Gabor kernel $\psi_{\mu,v}$ outputs $G_{\mu,v}(z) = Img(z) * \psi_{\mu,v}(z)$, where "*" denotes the convolution operator. The complex Gabor filtering coefficient $G_{\mu,v}(z)$ can be rewritten as

$$G_{\mu,v}(z) = M_{\mu,v}(z) \cdot \exp\left(i\theta_{\mu,v}(z)\right)$$

with $M_{\mu,v}$ being the magnitude and $\theta_{\mu,v}$ being the phase. It is known that magnitude information contain the variation of local energy in the image. As a multi-scale and multi-orientation feature extraction technique, Gabor filtering generates highly redundant features, and thus it is necessary to down-sample the filtering outputs to reduce the Gabor feature dimension as well as the time and space complexity. In [17], the augmented Gabor feature vector χ is defined via uniform down-sampling, normalization and concatenation of the Gabor filtering coefficients:

$$\chi = \left(a_{0,0}^{(\rho)}; a_{1,0}^{(\rho)}; \cdots; a_{7,4}^{(\rho)}\right)$$

where $a_{\mu,v}^{(\rho)}$ is the concatenated column vector of magnitude matrix $M_{\mu,v}^{(\rho)}$ down-sampled by a factor of ρ.

3. Gabor-feature based robust representation and classification

3.1. Gabor-feature based robust representation

Images from the same face, taken at (nearly) the same pose but under varying illumination, often lie in a low-dimensional linear subspace known as the *harmonic plane* or *illumination cone* [12][20]. This implies that if there are only variations of illumination, SRC can work very well. However, SRC with the holistic image features is less effective when there are local deformations of face images, such as certain amount of variations of expressions and pose.

The augmented Gabor face feature vector χ, which is a local feature descriptor, can not only enhance the face feature but also tolerate image local deformation to some extent. So we propose to use χ to replace the holistic face features for face representation, and the Gabor-feature based representation without face occlusion is

$$\chi(y_0) = X(A_1)\beta_1 + X(A_2)\beta_2 + \cdots + X(A_K)\beta_K = X(A)\beta \tag{6}$$

where $X(A) = \left[X(A_1)\, X(A_2) \cdots X(A_K) \right]$, $X(A_i) = \left[\chi(s_{i,1}), \chi(s_{i,2}), \cdots, \chi(s_{i,n_i}) \right]$, $\beta = [\beta_1; \beta_2; \cdots; \beta_K]$.

When the query face image is occluded, similar to SRC, an occlusion dictionary with Gabor features could be introduced to code the occlusion components, and the Gabor-feature based robust representation could be formulated as:

$$\chi(y) = \left[X(A), X(A_e) \right][\beta; \beta_e] = X(B)\omega \tag{7}$$

where $X(A_e)$ is the Gabor-feature based occlusion dictionary, and β_e is the coding vector of the input Gabor feature vector $\chi(y)$ over $X(A_e)$.

For the convenience of expression, we call the representation in either Eq. (6) (for FR without occlusion) or Eq. (7) (for FR with occlusion) the Gabor-feature based robust representation (GRR), and the representation vector in the GRR model can be solved by

$$\min_{\beta}\left\{ \left\| \chi(y_0) - X(A)\beta \right\|_2^2 + \lambda \left\| \beta \right\|_{l_p} \right\} \text{ or } \min_{\omega}\left\{ \left\| \chi(y) - X(B)\omega \right\|_2^2 + \lambda \left\| \omega \right\|_{l_p} \right\} \tag{8}$$

where $\left\| \bullet \right\|_{l_p}$ means the l_p-norm, and $p=1$ or 2 in our paper. In the case of occlusion, the selection of occlusion dictionary $X(A_e)$ has a big affect on the performance of GRR, and thus one key issue is how to define $X(A_e)$ to make the GRR effective and efficient.

3.2. Discussions on occlusion dictionary

SRC [10] is successful in solving the problem of face occlusion by introducing an occlusion dictionary A_e to code the occluded face components; however, one drawback of SRC is that the number of atoms in the used occlusion dictionary is very big. More specifically, the identity matrix was employed in SRC so that the number of atoms equals to the dimensionality of the image feature vector. For example, if the feature vector has a dimensionality of 3000, then the occlusion dictionary is of size 3000×3000. Such a high dimensional dictionary makes the sparse coding very expensive, and even computationally prohibitive. Suppose the size of the dictionary is $m \times n$, then the empirical complexity of the commonly used l_1-regularized sparse coding methods (such as l_1_ls [24], l_1_magic [25], and MOSEK [26]) to solve Eq. (1) is $O(m^2 n^\varepsilon)$ with $\varepsilon \approx 1.5$ [24][13]. So if the number of atoms (i.e., n) in the occlusion dictionary is too big, the computational cost will be huge, especially in dealing with FR with occlusion.
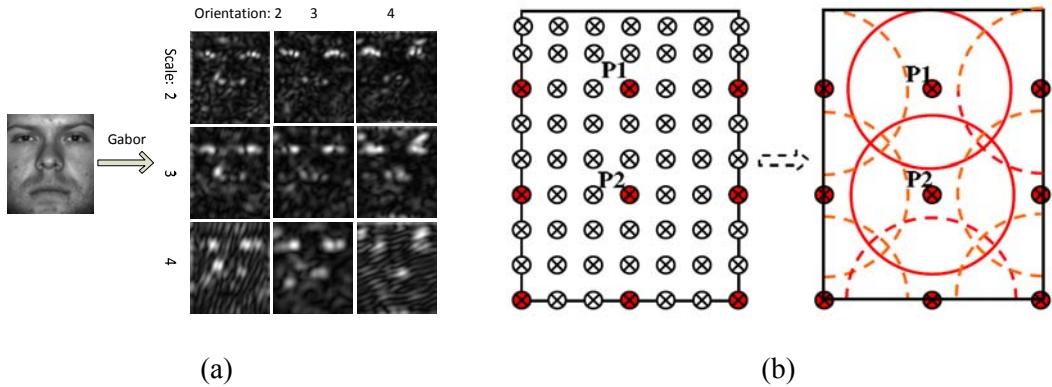

Figure 1: Gabor feature extraction. (a) Multi-scale and multi-orientation Gabor filtering. (b) The uniform down-sampling of Gabor filtering responses.

By using Gabor features for face representation, the feature dictionary A and the occlusion dictionary A_e in Eq. (3) will be transformed into the Gabor feature dictionary $X(A)$ and the Gabor-feature based occlusion dictionary $X(A_e)$ in Eq. (7). Fortunately, $X(A_e)$ is compressible. This can be easily illustrated by Fig. 1.

Fig. 1(a) illustrates the process of Gabor filtering. It is easy to see there are a rich amount of redundancies in the filtering responses across different scales and orientations. Therefore after the band-pass Gabor filtering, a uniform spatial down-sampling with a factor of ρ is conducted to form the augmented Gabor feature vector χ, as indicated by the red pixels in Fig. 1(b). The spatial down-sampling is performed for all the Gabor filtering outputs along different orientations and on different scales. Therefore, the number of (spatial) pixels in the augmented Gabor feature vector χ is $1/\rho$ times that of the original face image; meanwhile, at each location, e.g., P1 or P2 in Fig. 1(b), there is a set of directional and scale features extracted by Gabor filtering in the neighborhood (e.g., the circles centered on P1 and P2). Certainly, the directional and scale features at the same spatial location have some correlation, and there are often some overlaps between the supports of Gabor filters, which make the Gabor features at neighboring positions also have some redundancies.

Considering that "occlusion" is a phenomenon of spatial domain, a spatial down-sampling of the Gabor features with a factor of ρ implies that we can use approximately $1/\rho$ times the occlusion bases to code the Gabor features of the occluded face image. In other words, the Gabor-feature based occlusion dictionary $X(A_e)$ can be compressed because the Gabor features are redundant as we discussed above. To validate this conclusion, we suppose that the image size is 50×50, and in the original SRC the occlusion dictionary is an identity matrix $A_e = I \in \Re^{2500 \times 2500}$. Then the Gabor-feature based occlusion matrix $X(A_e) \in \Re^{2560 \times 2500}$, where the dimensionality of augmented Gabor feature is 2560 with $\rho=39.06$, $\mu=\{0,\ldots,7\}$, $v=\{0,\ldots,4\}$. Fig. 2 shows the singular values of $X(A_e)$. Obviously, although all the basis vectors of identity matrix I (i.e., A_e) have equal importance, only a few (60, with energy proportion of 99.67%) singular vectors of $X(A_e)$ have significant singular values, as shown in Fig. 2. This implies that $X(A_e)$ can be much more compactly represented by using only a few atoms generated from $X(A_e)$, often with a compression ratio about ρ:1. For example, in this experiment we have $2500/60=41.7 \approx \rho=39.06$. Next we present an algorithm to compute a more compact occlusion dictionary.
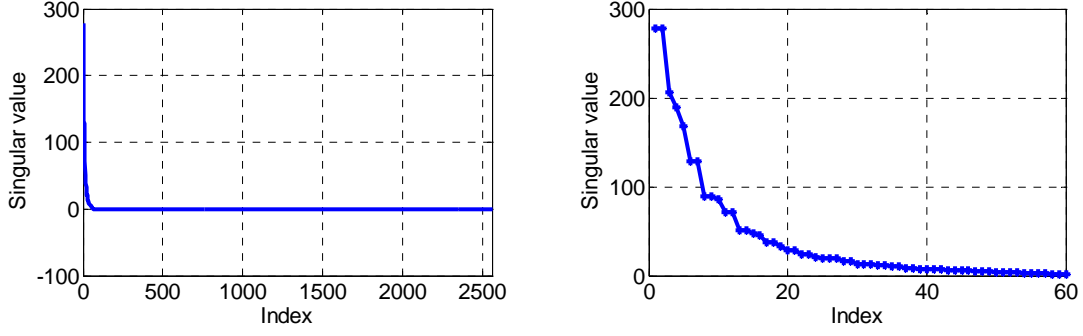
Figure 2: The singular values (left: all the singular values, right: the first 60 singular values) of Gabor feature-based occlusion matrix.

3.3. Gabor occlusion dictionary (GOD) computing

Now that $X(A_e)$ is compressible, we propose to compute a compact occlusion dictionary from it with suitable regularization on the coefficients. Here a compact dictionary, denoted by $D \in \Re^{m \times n}$, refers to a dictionary which has much less columns (i.e., the so-called atoms) than rows (i.e., $n \ll m$). From the view of handling occlusion, a compact occlusion dictionary means that the learnt dictionary atoms have lower correlation and stronger ability in representing face occlusions. In addition, the coding speed would be much faster because of the reduced size of occlusion dictionary. We call the computed compact occlusion dictionary Gabor occlusion dictionary (GOD) and denote it as Γ. Then we could replace $X(A_e)$ by Γ in the GRR based FR.

For the convenience of expression, we denote by $Z = X(A_e) = [z_1, \cdots, z_{n_e}] \in \Re^{m_\rho \times n_e}$ the original Gabor-feature based occlusion matrix, with each column z_i being the augmented Gabor-feature vector generated from each atom of A_e. The compact occlusion dictionary to be computed is denoted by $\Gamma = [d_1, d_2, ..., d_q] \in \Re^{m_\rho \times q}$, where q can be set as slightly less than n_e/ρ in practice. It is required that each occlusion basis d_j, j=1, 2, …, q, is a unit column vector, i.e. $d_j^T d_j = 1$. Since we want to replace Z by Γ, it is expected that the original dictionary Z can be well represented by Γ with the representation coefficients being regularized via l_p-norm regularization. Obviously, p=1 means that we require sparse representation on the learnt GOD. Inspired by the success of l_2-norm regularization in CRC [32], we can also use l_2-norm coefficient regularization. With such considerations, the objective function for determining Γ is defined as:

$$J_{\Gamma, A} = \arg\min_{\Gamma, A} \left\{ \|Z - \Gamma A\|_F^2 + \zeta \|A\|_{l_p} \right\} \quad \text{s.t.} \quad d_j^T d_j = 1, \forall j \tag{9}$$

11

where Λ is the representation matrix of $\mathbf{Z}$ over dictionary Γ, ζ is a positive scalar that balances the F-norm

term and the l_p-norm term (here p=1 for $\|\cdot\|_1$ and p=2 for $\|\cdot\|_F^2$).

Table 2: Algorithm of Gabor occlusion dictionary computing.

Algorithm of Gabor occlusion dictionary (GOD) computing

---

1. Initialize Γ
   We initialize each column of Γ as a random vector with unit l_2-norm.
2. Fix Γ and solve Λ
   By fixing Γ, the objective function in Eq. (9) will be reduced to

$$J_\Lambda = \arg\min_\Lambda \left\{ \|\mathbf{Z} - \Gamma\Lambda\|_F^2 + \zeta \|\Lambda\|_{l_p} \right\} \tag{10}$$

The minimization of Eq. (10) for p=1 can be achieved by the l_1-norm minimization techniques. In this paper, we use the algorithm in [24]. The minimization of Eq. (10) for p=2 could be efficiently solved since has a closed-form least square solution [32].

3. Fix Λ and update Γ
   Now the objective function is reduced to

$$J_\Gamma = \arg\min_\Gamma \left\{ \|\mathbf{Z} - \Gamma\Lambda\|_F^2 \right\} \quad \text{s.t.} \quad \mathbf{d}_j^T \mathbf{d}_j = 1, \forall j \tag{11}$$

We can write matrix Λ as $\Lambda=[\boldsymbol{\beta}_1;\boldsymbol{\beta}_2,\dots,\boldsymbol{\beta}_p]$, where $\boldsymbol{\beta}_j$, j=1,2,...,q, is the row vector of Λ. We update the occlusion bases one by one. When updating $\mathbf{d}_j$, all the other columns of Γ, i.e., $\mathbf{d}_l$, $l \neq j$, are fixed. Then J_Γ in Eq. (11) is converted into

$$J_{\mathbf{d}_j} = \arg\min_{\mathbf{d}_j} \left\| \mathbf{Z} - \sum_{l \neq j} \mathbf{d}_l \boldsymbol{\beta}_l - \mathbf{d}_j \boldsymbol{\beta}_j \right\|_F^2 \quad \text{s.t.} \quad \mathbf{d}_j^T \mathbf{d}_j = 1 \tag{12}$$

Let $\mathbf{Y} = \mathbf{Z} - \sum_{l \neq j} \mathbf{d}_l \boldsymbol{\beta}_l$, Eq. (12) can be written as

$$J_{\mathbf{d}_j} = \arg\min_{\mathbf{d}_j} \left\| \mathbf{Y} - \mathbf{d}_j \boldsymbol{\beta}_j \right\|_F^2 \quad \text{s.t.} \quad \mathbf{d}_j^T \mathbf{d}_j = 1 \tag{13}$$

Using Langrage multiplier, $J_{\mathbf{d}_j}$ is equivalent to

$$J_{\mathbf{d}_j,\gamma} = \arg\min_{\mathbf{d}_j} tr\left(-\mathbf{Y}\boldsymbol{\beta}_j^T \mathbf{d}_j^T - \mathbf{d}_j \cdot \boldsymbol{\beta}_j \mathbf{Y}^T + \mathbf{d}_j \cdot (\boldsymbol{\beta}_j \boldsymbol{\beta}_j^T - \gamma)\mathbf{d}_j^T + \gamma \right) \tag{14}$$

where γ is a scalar variable. Differentiating $J_{\mathbf{d}_j,\gamma}$ with respect to $\mathbf{d}_j$, and let it be **0**, we have

$$\mathbf{d}_j = \mathbf{Y}\boldsymbol{\beta}_j^T \left(\boldsymbol{\beta}_j \boldsymbol{\beta}_j^T - \gamma \right)^{-1} \tag{15}$$

Since $\left(\boldsymbol{\beta}_j \boldsymbol{\beta}_j^T - \gamma \right)$ is a scalar and γ is a variable, the solution of Eq. (15) under constrain $\mathbf{d}_j^T \mathbf{d}_j = 1$ is

$$\mathbf{d}_j = \mathbf{Y}\boldsymbol{\beta}_j^T / \left\| \mathbf{Y}\boldsymbol{\beta}_j^T \right\|_2 \tag{16}$$

Using the above procedures, we can update all the vectors $\mathbf{d}_j$, and hence the whole set Γ is updated.
4. Output Γ
   Go back to step 2 until the values of $J_{\Gamma,\Lambda}$ in adjacent iterations are close enough, or the maximum number of iterations is reached. Finally, output Γ.

Eq. (9) is a joint optimization problem of the occlusion dictionary Γ and the representation matrix Λ. Like in many multi-variable optimization problems, we solve Eq. (9) by optimizing Γ and Λ alternatively. The optimization procedures are described in the following Table 2. Based on our experiments, the random initialization of the dictionary affects little the learned GOD as well as the final face recognition accuracy. In general, about 10 iterations are needed to stop the algorithm of GOD.

It is straightforward that the above GOD computing algorithm converges because in each iteration $J_{\Gamma,\Lambda}$ will decrease. Fig. 3 illustrates an example of GOD on the AR database [21]. Based on our experiments, on other datasets the algorithm of GOD learning also converges quickly. Consequently, in GRR we use the GOD Γ to replace the $X(A_e)$ in Eq. (7). Finally, the coding problem in GRRC with face occlusion is

$$\min_{\boldsymbol{\omega}} \left\{ \left\| \boldsymbol{\chi}(\boldsymbol{y}) - \boldsymbol{B}_{\Gamma}\boldsymbol{\omega}_{\Gamma} \right\|_2^2 + \lambda \left\| \boldsymbol{\omega}_{\Gamma} \right\|_{l_p} \right\} \quad \text{where } \boldsymbol{B}_{\Gamma} = \left[X(A) \, \Gamma \right], \boldsymbol{\omega}_{\Gamma} = \left[\boldsymbol{\beta}; \boldsymbol{\beta}_{\Gamma} \right] \tag{17}$$
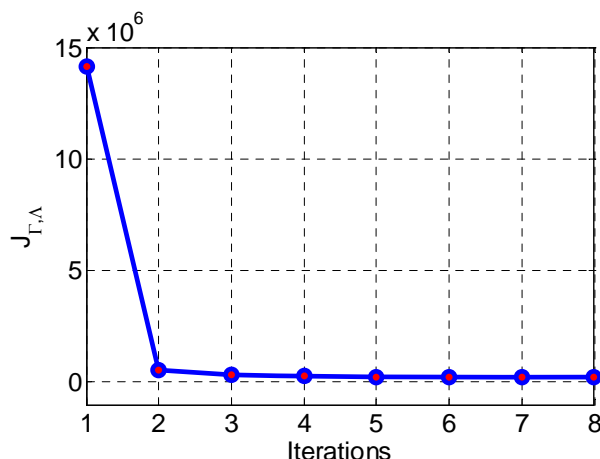

Figure 3: Illustration of the convergence of the proposed Gabor occlusion dictionary (GOD) computing algorithm on the AR database. A GOD with 100 atoms is computed from the original Gabor-feature based occlusion matrix with 4980 columns. The compression ratio is nearly 50:1.

3.4. GRR based classification (GRRC)

The SRC scheme [10] assumes that the face image representation residual is sparse, and thus uses the l_1-norm to characterize the representation coefficients associated with the occlusion dictionary, i.e., the identity matrix. Because the number of atoms in the identity matrix is very big (equal to the dimensionality of face image), it is necessary to impose the l_1-norm sparsity on the coding coefficients for a robust and unique representation, yet this makes the complexity of SRC very high. However, when Gabor feature is adopted, a

13

compact GOD $\boldsymbol{\Gamma}$ (with only about 1/40 times the size of the identity matrix) can be learnt, and thus it may

not be necessary to use the l_1-norm sparsity to regularize the coding coefficients over the dictionary anymore.

For a given face Gabor feature $\chi(\boldsymbol{y})$, often its dimensionality is much higher than the number of atoms in

dictionary $\boldsymbol{B_\Gamma}=[X(\boldsymbol{A})\,\boldsymbol{\Gamma}]$ after GOD computing, which means that the dictionary $\boldsymbol{B_\Gamma}$ is not over-complete, and

hence the system

$$\chi(\boldsymbol{y}) \approx \boldsymbol{B_\Gamma}\boldsymbol{\omega_\Gamma} \tag{18}$$

is generally an over-determined system. This implies that the solution of Eq. (18) is stable even without any

regularization. Although Eq. (18) is stable even without any regularization, a suitable the regularization (e.g.,

l_2-norm) could make the representation more stable. In addition, the regularization will make the

representation coefficients smaller, which would make the coefficients associated with wrong class have

lower value. This increases the discrimination of representation coefficients, which benefits the final face

recognition. In this paper we test the results by using both l_1-norm and l_2-norm to regularize the coding

coefficients. We name the GRR based classification (GRRC) with l_1-norm regularization GRRC_L$_1$, and the

GRRC with l_2-norm regularization GRRC_L$_2$. The GRRC algorithm is summarized in Table 3.

Table 3: Algorithm of GRR based Classification (GRRC).

Algorithm of GRRC
1. Input: Gabor feature dictionary $X(\boldsymbol{A})$, GOD $\boldsymbol{\Gamma}$, and the Gabor feature $\chi(\boldsymbol{y_o})$ (for testing sample without occlusion) or $\chi(\boldsymbol{y})$ (for testing sample with occlusion).

2. Solve the l_p-minimization ($p=1$ or 2) problem (the Lagrange formulation):

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \left\{ \left\| \chi(\boldsymbol{y_o}) - X(\boldsymbol{A})\boldsymbol{\beta} \right\|_2^2 + \lambda \left\| \boldsymbol{\beta} \right\|_{l_p} \right\} \tag{19}$$

or (let $\boldsymbol{\omega_\Gamma} = [\boldsymbol{\beta}; \boldsymbol{\beta_\Gamma}]$)

$$\hat{\boldsymbol{\omega}}_\Gamma = \arg\min_{\boldsymbol{\omega_\Gamma}} \left\{ \left\| \chi(\boldsymbol{y}) - [X(\boldsymbol{A})\,\boldsymbol{\Gamma}]\boldsymbol{\omega_\Gamma} \right\|_2^2 + \lambda \left\| \boldsymbol{\omega_\Gamma} \right\|_{l_p} \right\} \tag{20}$$

Where $\hat{\boldsymbol{\omega}}_\Gamma = [\hat{\boldsymbol{\beta}}; \hat{\boldsymbol{\beta}}_\Gamma]$ and λ is a positive scalar that balances the coding residual and regularization strength.

3. Compute the residuals

$$r_i(\boldsymbol{y_o}) = \left\| \chi(\boldsymbol{y_o}) - X(\boldsymbol{A}_i)\delta_i(\hat{\boldsymbol{\beta}}) \right\|_2, \quad \text{for } i = 1, \cdots, K. \tag{21}$$

or

$$r_i(\boldsymbol{y}) = \left\| \chi(\boldsymbol{y}) - \boldsymbol{\Gamma}\hat{\boldsymbol{\beta}}_\Gamma - X(\boldsymbol{A}_i)\delta_i(\hat{\boldsymbol{\beta}}) \right\|_2, \quad \text{for } i = 1, \cdots, K. \tag{22}$$

4. Output identity($\boldsymbol{y_o}$)=argmin$_i$ $r_i(\boldsymbol{y_o})$ or identity ($\boldsymbol{y}$)=argmin$_i$ $r_i(\boldsymbol{y})$.

3.5. Time complexity

The empirical complexity of the commonly used l_1-regularized sparse coding methods is $O(m^2 n^\varepsilon)$ with $\varepsilon \approx$ 1.5 [24][13], while the time complexity of l_2-norm regularized coding is only $O(mn)$ [32] for that the coding projection matrix could be computed offline, where m is facial feature dimensionality and n is the number of dictionary atoms. For GRRC, in Fourier domain it is very fast to extract Gabor features, whose time complexity could be negligible compared with that of l_1-norm regularized sparse coding.

In the case of FR without occlusion, n is the number of training samples. Therefore, GRRC_L_1 has similar computational burden to SRC, but GRRC_L_2 has much lower time complexity than GRRC_L_1 and SRC. For FR without occlusion, there is a fast version of SRC, namely SRC using Hashing [46]. This method is usually faster than the original SRC because the used random projection matrix is very sparse. So GRRC_L_1 would have a little higher time complexity than SRC using Hashing, but GRRC_L_2 is still much faster than SRC using Hashing.

In the case of FR with occlusion, it is easy to get that the time complexity of GRRC_L_1 is $O(m^2(n+m/\rho)^\varepsilon)$, where $\rho \approx 40$. This is much lower than SRC whose time complexity is $O(m^2(n+m)^\varepsilon)$. Obviously, GRRC_L_2's time complexity is $O(m(n+m/\rho))$ and it is the fastest one among the three methods.

4. Experimental results

In this section, we present experiments on benchmark face databases to demonstrate the superiority of GRRC to SRC. Before giving the detailed experimental results, we discuss the selection of Gabor features and regularization of GOD computing in Section 4.1. To evaluate more comprehensively the performance of GRRC, in Section 4.2 we first test FR with little deformation; then in Section 4.3 we demonstrate the robustness of GRRC to expression and pose variation; finally in Section 4.3 we test FR against block occlusion and real disguise. In our implementation of Gabor filters, the parameters are set as $K_{max}=\pi/2$, $f = \sqrt{2}$, $\sigma=1.5\pi$, $\mu=\{0,\ldots,7\}$, $v=\{0,\ldots,4\}$ by our experiences and they are fixed for all the experiments. The parameter λ in GRRC should be set as a small positive value to make the representation more stable and the

15

coding coefficient regularized. A large value of λ would make the regularization too strong so that the signal representation fidelity can be reduced, resulting in the decrease of recognition accuracy. In the experiments, λ in GRRC is fixed as 0.0005 for FR without and with occlusion. We also give the results of GRRC with $\lambda=0.001$ for FR without occlusion to show that GRRC is very robust to parameter's value. In addition, all the face images are cropped and aligned by using the location of eyes, which is provided by the face databases (for Mulit-PIE, we manually locates the positions of eyes).

In the following tables of this section, the results of competing methods with reference numbers are cropped from the original papers. All the other results are computed by us with reporting their best recognition rates.

4.1. Gabor features and regularization of GOD computing

1) Gabor features: In GRRC, we adopt the Gabor magnitude as the augmented facial features. Here we also evaluate other Gabor features, such as Gabor real parts, Gabor imaginary parts, and the concatenation of Gabor real and imaginary parts. We replace Gabor magnitude features in GRRC_L_2 by these Gabor features, and test their performance on the AR database (the detailed experimental setting is described in Section 4.2). Table 4 lists the recognition rates. It is easy to see that the features of Gabor real parts (denoted by GRRC_L_2 (Real parts)), Gabor imaginary parts (denoted by GRRC_L_2 (Imaginary parts)) and their concatenation (denoted by GRRC_L_2 (Real + Imaginary)) do not lead to good results. This demonstrates that Gabor magnitude (denoted by GRRC_L_2 (Magnitude)) is more discriminative in the Gabor feature-based representation scheme. The results by SRC [10] and CRC [32] schemes with holistic PCA features are also listed in the Table for comparison.

Table 4: Face recognition rates (%) with different Gabor features on AR database.

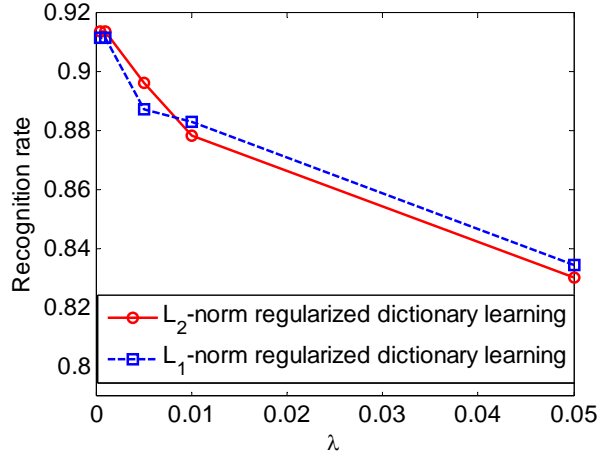| Dimension	130	300	540
PCA+SRC	89.7	93.3	93.5
PCA+CRC	90.0	93.7	93.9
GRRC_L_2 (Real parts)	84.3	89.4	91.4
GRRC_L_2 (Imaginary parts)	85.8	91.0	93.3
GRRC_L_2 (Real + Imaginary)	85.0	91.4	93.6
GRRC_L_2 (Magnitude)	**93.1**	**96.8**	**97.3**

Figure 4: Recognition rates by using l_1-norm and l_2-norm regularized GOD computing in the experiment of FR with random block occlusion.

2) Regularization on GOD computing: In the GOD computing algorithm (refer to Table 2), we regularize the coding coefficient by l_p-norm with p=1 or 2. Here we use an FR experiment on Extended Yale B[19][20] with random block face occlusion (about 45% occlusion) to discuss the selection of l_p-norm. The detailed experimental setting will be presented in the experiments of *FR with random block occlusion* in Section 4.4. We set the parameter ζ in the model (Eq. (9)) of GOD computing as 0.005. The recognition rates of l_p-norm regularized GOD computing versus different regularization parameters λ in coding (Eq. (20) with l_1-norm regularization) of the classification stage are shown in Fig. 4. It can be seen that there is not much difference in recognition accuracy between l_1-norm and l_2-norm regularization in GOD computing. The reason is that the redundancy of Gabor feature transformation (analyzed in Section 3.2) makes the learnt GOD dictionary compact so that the GOD dictionary is obviously over-determined. An over-determined dictionary itself could stably represent the testing sample even without regularization, while the l_1-norm or l_2-norm constraint on coding coefficient in GRRC could make the representation more stable and make the coding coefficients more discriminative. Therefore, the l_1-norm and l_2-norm regularizations will lead to stable occluded face representation and similar recognition results. Considering that the recognition rates by l_1-norm and l_2-norm regularized GOD computing are similar, we prefer to use the l_2-norm regularized one for its fast speed. In our paper, the parameter ζ in GOD computing is set as a small scalar, e.g., 0.001.

In order to give an intuitive illustration of the leant GOD, we plot the 1st, 51st, 101st and 151st atom of l_1-norm regularized GOD in Fig. 5. We could see that the learnt GOD atoms are roughly periodic signals,

which have 40 repeated patterns, and each pattern corresponds to one orientation on one scale of the Gabor feature (the Gabor feature is the concatenation of 40 down-sampling Gabor magnitudes). The original occlusion dictionary (i.e., the identity matrix) has clear spatial meaning, e.g., each atom is a unit vector representing one pixel of the image. However, the size of such an occlusion dictionary is too big (e.g., 8064×8064 in this experiment). Because the spatial size of the augmented down-sampling Gabor feature is greatly reduced and occlusion is a phenomenon in spatial domain, the number of atoms in GOD could be greatly reduced. The learnt GOD not only has much smaller size (e.g., 8940×200), but also have very clear spatial meaning, i.e., on each down-sampled Gabor magnitude feature, the corresponding atom of GOD is a local basis to represent the scale and orientation information at that location. Therefore, GOD is much more efficient to handle occlusion.
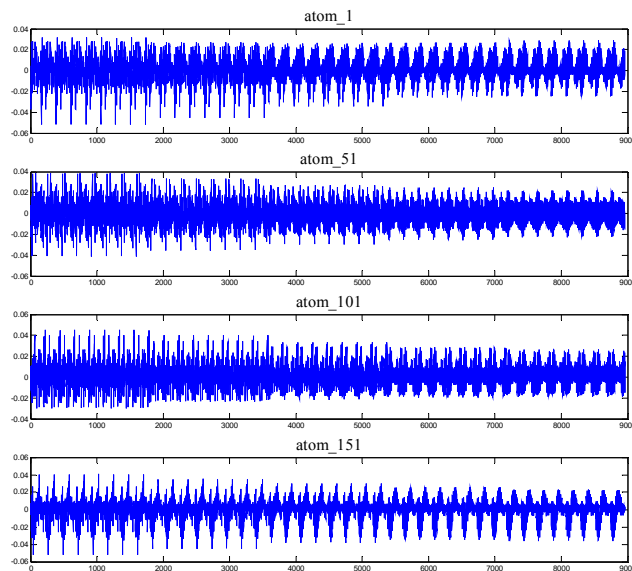
Figure 5: The 1st, 51st, 101st, and 151st atoms of the learnt Gabor Occlusion Dictionary.

4.2. Face recognition with little deformation

We evaluate the proposed GRRC scheme on four representative facial image databases: Extended Yale B [19][20], AR [21], Multi-PIE [27] and FERET[22][23]. We compare GRRC with SRC [10], CRC [32], Linear Regression for Classification (LRC) [37], linear Support Vector Machine (SVM) and Nearest Neighbor (NN) methods. If no specific instruction, for all the competing methods we use PCA to reduce the feature dimension.

1) Extended Yale B Database: The Extended Yale B database consists of 2,414 frontal-face images of 38 individuals, captured under various laboratory-controlled lighting conditions [19][20]. For each subject, we randomly selected half of the images for training (i.e., 32 images per subject), and used the other half for testing. The images are normalized to 192×168, and the dimension of the augmented Gabor feature vector of each image is 19760 (40×26×19). The results of all the methods versus the feature dimension are listed in Table 5. It can be seen that GRRC is better than SRC, CRC and other methods in all the dimensions except that SRC is slightly better GRRC_L_2 in the dimension of 56. This shows that the l_1-norm sparse constraint will be more effective than the non-sparse l_2-norm constraint in classification when the discrimination of feature (e.g., 56-dimensional) is not high and the dictionary (e.g., with the size of 56×1206) is an over-complete matrix. GRRC_L_2 has similar performance to GRRC_L_1 when the dimension is greater than 56. On this database, the maximal recognition rates of the competing methods are 99.2% for GRRC_L_1, 99.1% for GRRC_L_2, 97.9% for SRC, 98.0% for CRC, 96.4 for SVM, 95.7% for LRC, and 92.0% for NN. In addition, it can be seen that GRRC is not sensitive to the value of λ.

Table 5: Face recognition results (%) on the Extended Yale B database. For GRRC, r_1 (r_2) means that r_1 is the recognition rate for λ=0.0005, and r_2 for λ=0.001.

	56	120	300	504
SRC	92.6	95.6	97.4	97.9
CRC	88.6	95.4	97.4	98.0
NN	81.4	89.2	91.9	92.0
LRC	94.1	94.7	95.4	95.7
SVM	92.6	95.3	96.3	96.4
GRRC_L_1	**92.7(92.7)**	95.6(96.2)	97.9(97.9)	99.0(**99.2**)
GRRC_L_2	90.5(90.5)	**96.3(96.3)**	**98.4(98.4)**	99.1(99.1)

2) AR database: The AR database consists of over 4,000 frontal images from 126 individuals [21]. For each individual, 26 pictures were taken in two separate sessions. As in [10], in the experiment we chose a subset of the dataset consisting of 50 male subjects and 50 female subjects. For each subject, the seven images with illumination change and expressions from Session 1 were used for training, and the other seven images with only illumination change and expression from Session 2 were used for testing. The size of original face image is 165×120, and the Gabor-feature vector is of dimension 12000 (40×20×15). The comparison of GRRC and the competitors are shown in Table 6. Again we can see that GRRC performs much better than all the other methods under all the dimensions, especially with more than 3% improvement

when the dimension is lager than 54. On this database, the maximal recognition rate of GRRC_L_1, GRRC_L_2, SRC, CRC, SVM are 97.1%, 97.3%, 93.5%, 93.9% and 88.8%, respectively.

Table 6: Face recognition results (%) on the AR database. For GRRC, r_1 (r_2) means that r_1 is the recognition rate for λ=0.0005, and r_2 for λ=0.001.

	54	130	300	540
SRC	80.0	89.7	93.3	93.5
CRC	80.5	90.0	93.7	93.9
NN	67.8	70.1	71.2	72.1
LRC	75.4	76.0	70.7	76.7
SVM	77.5	82.7	87.3	88.8
GRRC_L_1	**86.0(86.0)**	**94.0(94.0)**	**96.7**(96.6)	97.1(97.1)
GRRC_L_2	82.7(82.7)	93.1(93.1)	**96.7(96.7)**	**97.3(97.3)**

Table 7: Face recognition results (%) on the Multi-PIE database. For GRRC, r_1 (r_2) means that r_1 is the recognition rate for λ=0.0005, and r_2 for λ=0.001.

	SRC	CRC	NN	LRC	SVM	**GRRC_L_1**	**GRRC_L_2**
Session 2	93.9	94.1	86.4	87.1	85.2	97.3(**97.5**)	97.1(97.2)
Session 3	90.0	89.3	78.8	81.9	78.1	96.7(96.7)	**96.8(96.8)**
Session 4	94.0	93.3	82.3	84.3	82.1	98.6(98.6)	**98.7(98.7)**

Table 8: Face recognition results (%) on the FERET database. For GRRC, r_1 (r_2) means that r_1 is the recognition rate for λ=0.0005, and r_2 for λ=0.001.

	SRC	CRC	NN	SVM	**GRRC_L_1**	**GRRC_L_2**
Fb	86.9	85.4	87.1	87.1	**95.7**(95.6)	95.6(95.6)
Fc	77.3	75.8	73.2	73.2	**97.4(97.4)**	94.8(95.4)
Dup1	51.6	51.5	47.8	47.8	77.7(78.0)	**79.1**(78.9)
Dup2	33.3	35.5	23.9	23.9	75.6(76.5)	**78.6(78.6)**

3) Large-scale Multi-PIE database: The CMU Multi-PIE database [27] contains images of 337 subjects captured in four sessions with simultaneous variations in pose, expression, and illumination. In the experiments, all the 249 subjects in Session 1 were used. For the training set, we used the 14 frontal images with illuminations {0,1,3,4,6,7,8,11,13,14,16,17,18,19} and neutral expression. For the testing sets, 10 typical frontal images of even-number illuminations taken with neutral expressions from Session 2 to Session 4 were used. The image size is cropped and normalized to 100×82, and the Gabor feature vector is of the dimension of 8320 (40×16×13). We use PCA to reduce the dimensionality of the input feature to 300. Table 7 lists the recognition rates in three tests by the competing methods. The results validate that GRRC methods get the best in accuracy, at least 3% higher than that of SRC and CRC in session 2 and about 5%

higher than that of SRC and CRC in other sessions. NN, LRC and SVM can not get good recognition accuracy (lower than 90%) in this database, much lower than SRC, CRC and GRRC.

4) Large-scale FERET database: The FERET database [22][23] is often used to validate an algorithm's effectiveness because it contains many kinds of image variations. By taking 'Fa' subset as a gallery, the probe subsets 'Fb' and 'Fc' were captured with expression and illumination variations. Especially, 'Dup1' and 'Dup2' consist of images that were taken at different times with more than one year interval. Here we should note that in the Gallery set 'Fa', each subject only has one sample, which is very challenging for SRC and GRRC because usually they usually need several samples for each subject to construct the subspace. The image size is cropped and normalized to 150×130, and the Gabor feature vector is of dimension 21000 ($40 \times 25 \times 21$). For all the competing methods, we used LDA to reduce the original feature dimensionality to 428 for LDA could achieve better performance than PCA in this challenging dataset. Table 8 shows the face recognition results on FERET database. It is surprised that SRC and CRC have higher accuracy than NN and SVM except for 'Fb' even only one sample for each subject in the training set. GRRC methods achieve the best performance with over 95% recognition rates in 'Fb' and 'Fc' and about 78% in 'Dup1' and 'Dup2'. It can also be seen that for 'Fb', GRRC has at least 8% improvements compared to other methods, while with about 20%, 27% and 43% improvements for 'Fc', 'Dup1' and 'Dup2', respectively. According to the recent state-of-the-art FR results on the FERET database, e.g., Xie *et al.* 's method [45], further improvement could be achieved if more discriminative features, e.g., fused Gabor magnitude and phase feature [45], are utilized in the framework of GRRC.

From the experimental results in Extended Yale B, AR, Multi-PIE and FERET, we could see that GRRC is very robust to the value of λ and GRRC_L$_1$ and GRRC_L$_2$ have very similar performance (the gap usually is less than 0.5% in high dimensional feature), showing that GRRC_L$_2$ is very suitable for the practical FR systems due to its fast speed and good performance. Besides, the improvements brought by GRRC on the AR, Multi-PIE, and FERET are much bigger than that on the Extended Yale B database. This is because mostly there is only illumination variation between the training images and testing images, and the number of training samples (i.e., 32) in the Extended Yale B database is also high. Thus the original SRC and CRC work well on it. However for the more challenging cases (e.g., the training and testing samples of the AR,

Multi-PIE and FERET have much more variations, including time, illumination, etc., but with very limited number of training samples), the local feature based GRRC is much more robust than the holistic feature based SRC, CRC, SVM, LRC and NN in this case.

4.3. Face recognition with pose and expression variations

In this section, we verify the robustness of GRRC to pose and expression variations on the pose subset of FERET database [22][23] and expression subset of Multi-PIE [27].

1) FERET pose database: Here we used the pose subset of the FERET database [22][23], which includes 1400 images from 198 subjects (about 7 each). This subset is composed of the images marked with '*ba*', '*bd*', '*be*', '*bf*', '*bg*', '*bj*', and '*bk*'. In our experiment, each image has the size of 80×80 and the dimension of Gabor feature is 6760 (40×13×13). Some sample images of one person are shown in Fig. 6.

Five tests with different pose angles were performed. In test 1 (pose angle is zero degree), images marked with '*ba*' and '*bj*' were used as the training set, and images marked with '*bk*' were used as the testing set. In all the other four tests, we used images marked with '*ba*', '*bj*' and '*bk*' as gallery, and used the images with '*bg*', '*bf*', '*be*' and '*bd*' as probes, respectively. Here we use 350-dimension Eigenfaces as the input feature. Table 9 lists the results of different methods for various face poses. Obviously, we can see that GRRC has much higher recognition rates than SRC and other methods. In particular, when the pose variation is moderate (0° and ±15°), about 20% improvement is achieved by GRRC compared to SRC. We could also see that GRRC_L_2 performs very similarly to GRRC_L_1. It is undeniable that GRRC's performance also degrades much when pose variation becomes large (e.g. ±25°). Nevertheless, GRRC can much improve the robustness to moderate pose variation, and thus it could tolerate registration error (e.g., pose variation, misalignment) to some extent.

| ba: gallery | bj: expression | bk: illumination | be: +15 | bf: -15 | bg: -25 | bd: +25 |

Figure 6: Some samples of a subject on the pose subset of the FERET database.

Table 9: Face recognition results (%) on the pose subset of the FERET database. For GRRC, r_1 (r_2) means that r_1 is the recognition rate for λ=0.0005, and r_2 for λ=0.001.

Pose (degree)	-25	-15	0	15	25
SRC	32.5	70.5	83.5	57.5	28.0
CRC	21.0	62.5	74.5	40.0	17.0
NN	10.5	54.0	78.5	39.0	17.5
LRC	11.5	58.0	75.5	40.5	20.0
SVM	14.5	61.5	80.5	43.5	20.0
GRRC_L₁	41.5**(42.0)**	**95.5(95.5)**	**99.0(99.0)**	89.0(89.5)	44.5**(45.0)**
GRRC_L₂	41.0(41.0)	**95.5(95.5)**	**99.0(99.0)**	91.5(91.5)	44.0(44.0)

2) Multi-PIE expression subset: All the 249 subjects in Session 1 were used as training set in this experiment. To make the FR more challenging, four subsets with both illumination and expression variations in Sessions 1, 2 and 3 were used for testing. For the training set, as in [34] we used the 7 frontal images with extreme illuminations {0, 1, 7, 13, 14, 16, 18} and neutral expression (refer to Fig. 7(a) for examples). For the testing set, 4 typical frontal images with illuminations {0, 2, 7, 13} and different expressions (smile in Sessions 1 and 3, squint and surprise in Session 2) are used (refer to Fig. 7(b) for examples with surprise in Session 2, Fig. 7(c) for examples with squint in Session 2, Fig. 7(d) for examples with smile in Session 1, and Fig. 7(e) for examples with smile in Session 3). We used the Eigenface with dimensionality 900 as the face feature.
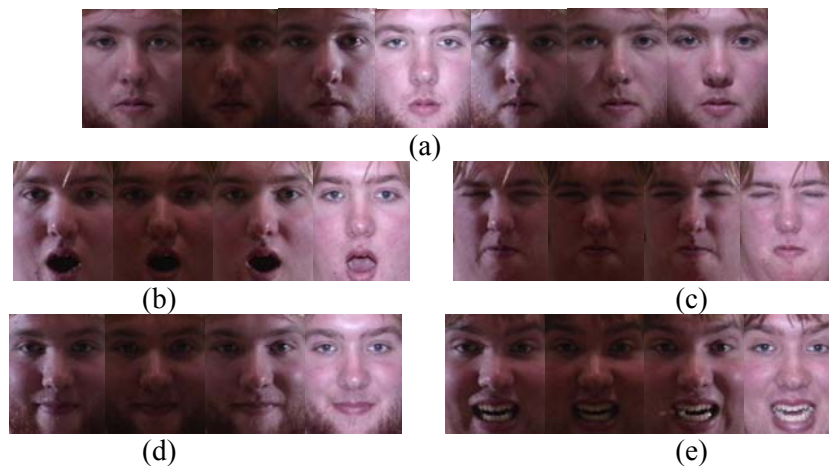

(a)

(b)

(c)

(d)

(e)

Figure 7: A subject in Multi-PIE database. (a) Training samples with only illumination variations. (b) Testing samples with surprise expression and illuminations in Session 2. (c) Testing samples with squint expression and illuminations in Session 2. (d) and (e) show the testing samples with smile expression and illumination variations in Session 1 and Session 3, respectively.

Table 10: Face recognition rates on Multi-PIE expression database. For GRRC, $r_1 (r_2)$ means that r_1 is the recognition rate for λ=0.0005, and r_2 for λ=0.001.

	Smile-S1	Smile-S3	Surprise-S2	Squint-S2
SRC	94.1	60.9	55.0	57.2
CRC	92.4	56.7	49.2	52.7
NN	89.4	46.3	40.5	50.3
LRC	90.4	49.8	40.1	52.1
SVM	88.9	46.3	25.6	47.7
Hash+OMP	92.2	50.2	42.3	51.8
Hash+L_1	87.2	50.0	46.4	56.2
GRRC_L_1	**97.7**(97.4)	73.4(73.3)	81.8**(82.4)**	87.5(87.8)
GRRC_L_2	97.3(97.3)	**74.2(74.2)**	82.2(82.2)	88.0**(88.1)**

Table 11: Average time (second) comparison for sparse representation-based FR methods

Method	SRC	Hash+OMP	Hash+L_1	GRRC_L_1	GRRC_L_2
Running Time	1.398	1.061	2.644	0.2423+1.400	0.2423+0.046

Table 10 lists the recognition rates in four testing sets by the competing methods, including SRC using Hasing [46] (e.g., Hash+OMP and Hash+L_1). It can be seen that GRRC achieves the best performance in all tests and SRC performs the second best. It can also be seen that SRC using Hashing has lower recognition rates than SRC, which may result from its use of random projection matrix for dimensionality reduction. In addition, all the methods achieve their best results when Smile-S1 is used for testing because the training set is also from Session 1. The highest rates of GRRC_L_1 and GRRC_L_2 are 97.4% and 97.3%, respectively, more than 3% improvement over the third best one, SRC. From testing set Smile-S1 to set Smile-S3, the variations increase because of the longer data acquisition time interval and expression changes (refer to Fig. 7 (d) and Fig. 7 (e)). The recognition rates of GRRC_L_1 and GRRC_L_2 drop by 24.1% and 23.1%, respectively, while those of SRC, CRC, NN, LRC and SVM drop by 33.2%, 35.7%, 43.1%, 40.6% and 42.6%, respectively, which validates that GRRC is much more robust to face variation than the other methods. For the testing set of Surpise-S2 and Squint-S2, GRRC has about 30% improvement over all the other methods. Meanwhile, for all the four tests, GRRC with l_1-norm constraint or l_2-norm constraint on coding coefficients has similar performance.

The running time of GRRC, SRC, and SRC using Hashing [46] (e.g., Hash+OMP and Hash+L_1) is compared in Table 11. Here the Gabor feature extraction for GRRC is 0.2423 second. From Table 11, it is

very clear that GRRC_L_2 is the fastest one, about 4 times faster than Hash+OMP, the second fastest method. GRRC_L_1's running time is still lower than Hash+L_1.

From the experiments on FR with local deformation (e.g., pose and expression variations), we could see that there is almost no difference between GRRC with λ=0.001 and GRRC with λ=0.0005, showing that GRRC is very robust to the value of λ. GRRC is much superior to the other methods, including SRC and CRC. This not only shows that collaborative representation based classification strategy with l_1 or l_2 norm regularization is more powerful than other classifiers, such as NN, LRC and SVM, but also demonstrates that Gabor magnitude features are more robust to the variations of pose and expression.

4.4. Recognition against occlusion

In this sub-section, we test the robustness of GRRC to face occlusions, including block occlusion and real disguise. FR with random block occlusion is performed on the Extended Yale B database [19][20], while FR with real disguise is performed on the AR database [21].

1) FR with random block occlusion: As in [10], we chose Subsets 1 and 2 (717 images, normal-to-moderate lighting conditions) for training, and Subset 3 (453 images, more extreme lighting conditions) for testing. In accordance to the experiments in [10], the images were resized to 96×84, and the occlusion dictionary A_e in SRC is set to an identity matrix.

With the above settings, in SRC the size of matrix $\boldsymbol{B}$ in Eq. (3) is 8064×8761. In the proposed GRRC, the dimension of augmented Gabor-feature vector is 8960 (40×16×14, ρ≈40). The GOD $\boldsymbol{\Gamma}$ is then computed using Algorithm in Table 2. In the experiment, we set the number of atoms in $\boldsymbol{\Gamma}$ to 200 (i.e., p=200, with compression ratio about 40:1), and hence the size of dictionary $\boldsymbol{B_\Gamma}$ in Eq. (17) is 8960×917. Compared with the original SRC, the dictionary size of GRRC is reduced from 8761 to 917.
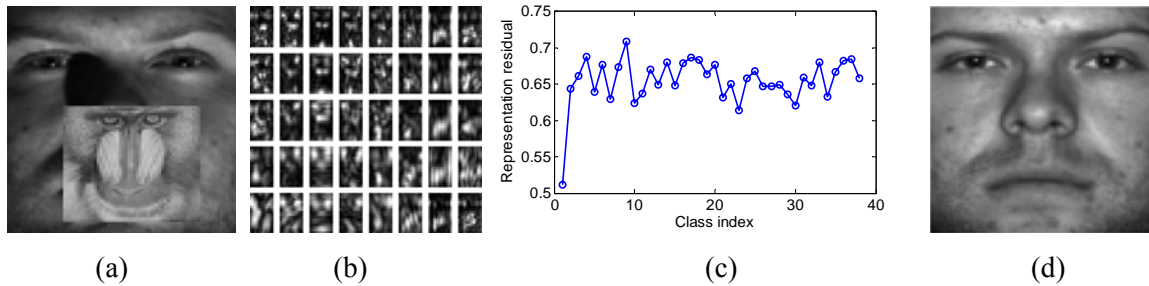
Figure 8: An example of face recognition with block occlusion. (a) A 30% occluded test face image y from the first class of Extended Yale B. (b). Uniformly down-sampled Gabor features $\chi(y)$ of the test image. (c). Estimated residuals $r_i(y)$, $i = 1, 2, …, 38$. (d). One sample of the class to which the test image is classified.

Table 12: The recognition rates (%) of different methods under different levels of block occlusion.

Occlusion ratio	0%	10%	20%	30%	40%	50%
SRC[10]	**100**	**100**	99.8	98.5	90.3	65.3
CRC	**100**	99.8	96.7	86.3	74.8	61.0
PCA+NN[10]	92.5	90.7	84.0	73.5	61.5	45.0
GRRC_L$_1$	**100**	**100**	**100**	**100**	96.5	**87.4**
GRRC_L$_2$	**100**	**100**	**100**	**100**	**97.1**	84.1

As in [10], we simulated various levels of contiguous occlusion, form 0% to 50%, by replacing a randomly located square block in each test image with an unrelated image, whose size is determined by the occlusion percentage. The location of occlusion was randomly chosen for each test image and is unknown to the computer. Fig. 8 illustrates the classification process by using an example. Fig. 8 (a) shows a test image with 30% randomly located occlusion; Fig. 8 (b) shows the augmented Gabor features of the test image. The residual of GRRC_L$_2$ are plotted in Fig. 8(c), and a template image of the identified subject is shown in Fig. 8(d). The detailed recognition rates of GRRC, SRC, CRC and PCA+NN (used as the baseline) are listed in the Table 12. We see that GRRC can correctly classify all the test images when the occlusion percentage is less than or equal to 30%. When the occlusion percentage becomes larger, the advantage of GRRC over SRC is getting higher. Especially, GRRC_L$_1$ can still have a recognition rate of 87.4% when half of image is occluded, while SRC and CRC only achieve a rate of 65.3% and 61.0% respectively. PCA+NN gets the worst results for it does not consider the occlusion. We could also see that good performance is still achieved when the representation coefficients on Gabor occlusion dictionary are regularized by l_2-norn in GRRC_L$_2$.

2) FR with real disguise: A subset from the AR database was used in this experiment. This subset consists of 1199 images from 100 subjects (14 samples each class except for a corrupted image w-027-14.bmp), 50 male and 50 female. 799 images (about 8 samples per subject) of non-occluded frontal views

with various facial expressions were used for training, while the others for testing. The images are resized to 83×60. So in original SRC, the size of matrix $\boldsymbol{B}$ in Eq. (3) is 4980×5779. In the proposed GRRC, the dimension of Gabor-feature vectors is 5200 (40×13×10, $\rho \approx 38$), and 100 atoms (with compression ratio 50:1) are computed to form the GOD by Algorithm in Table 2. Thus the size of dictionary $\boldsymbol{B}_\Gamma$ in Eq. (17) is 5200×899, and the dictionary size is reduced from 5779 to 899 for GRRC.

We consider two separate test sets of 200 images (1 sample each session and each subject, with neural expression). The first test set contains images of the subjects wearing sunglasses, which occlude roughly 20% of the image. The second test set is composed of images of the subjects wearing a scarf, which occlude roughly 40% of the images. The results by GRRC, SRC, CRC, PCA+NN and SVM are listed in Table 13 (where the results of SRC and PCA+NN are copied from the original paper [10]). We see that on faces occluded by sunglasses, GRRC achieves a recognition rate of 93.0%, over 5% higher than that of SRC, while for occlusion by scarves, the proposed GRRC_L$_1$ (GRRC_L$_2$) achieves a recognition rate of 79% (77.5%), about 20% higher than that of SRC. It is surprising that CRC gets 90.5% in the scarf case but with very low recognition accuracy in sunglass case. SVM gets bad performance for that it cannot learn the occlusion information from the training set without occlusion.

In [10], the authors also partitioned the image into blocks for face classification by assuming that the occlusion is continuous. Such an SRC scheme is denoted by SRC-p, with the CRC scheme denoted by CRC-p. Here, after partitioning the image into several blocks, we calculate the Gabor features of each block and then use GRRC to classify each block image. The final classification result is obtained by voting. We denote by GRRC-p the GRRC with partitioning. In experiments, as [10] we partitioned the images into eight (4×2) blocks of size 20×30. The Gabor-feature vector of each block is of dimension 800, and the number of atoms in the computed GOD $\boldsymbol{\Gamma}$ is set to 20. Thus the dictionary $\boldsymbol{B}$ in SRC is of size 600×1379, while the dictionary $\boldsymbol{B}_\Gamma$ in GRRC is of size 800×819. The recognition rates of SRC-p, CRC-p and GRRC-p are also listed in Table 13. We see that with partitioning, GRRC can lead to recognition rates of 100% on sunglasses and 99% on scarves, also better than SRC and CRC.

Table 13: Recognition rates (%) on the AR database with disguise occlusion ('-p': partitioned, '-sg': sunglasses, and '-sc': scarves).

	Sunglass		Scarf	
SRC (SRC-p) [10]	87	(97.5)	59.5	(93.5)
CRC (CRC-p) [32]	68.5	(91.5)	**90.5**	(95)
PCA+NN [10]	70.0		12.0	
SVM	66.5		16.5	
GRRC_L$_1$ (GRRC-p_L$_2$)	**93.0**	**(100)**	**79.0**	**(99)**
GRRC_L$_2$ (GRRC-p_L$_2$)	**93.0**	**(100)**	77.5	**(99)**

3) Running time comparison: The recognition rates and running time of the proposed GRRC and SRC on a more challenging FR experiment with real disguise are compared here. A subset of 50 males and 50 females are selected from the AR database. For each subject, 7 samples with no occlusion from session 1 are used for training, with all the remaining samples with disguises for testing. These testing samples (including 3 sunglass samples in Session1, 3 sunglass samples in Session 2, 3 scarf samples in Session 1 and 3 scarf samples in Session 2 per subject) not only have disguises, but also have variations of time and illumination. The image size and the extraction of Gabor feature of GRRC remains the same as before. Here λ=0.005 for GRRC and the programming environment is Matlab version R2011a. The desktop used is of 1.86 GHz CPU and with 2.99G RAM. All the l_1-minimization problem is solved by using the fast solver: ALM [43][44]. The recognition rates and running time of GRRC and SRC are listed in Table 14. The recognition rates of GRRC in all cases are much higher than SRC and CRC, especially with over 7% improvement on FR with sunglasses of session 1, and at least 43% in FR with scarf. It can also be seen that GRRC_L$_1$ is slightly better in FR with scarf, while GRRC_L$_2$ slightly better in FR with sunglasses. Fig. 9 plots the representation coefficients and residuals of a sample from class 1. As shown in Fig. 9(b), the sample is wrongly classified by GRRC_L$_1$ though the coefficients are sparse (see Fig. 9(a)). Although the representation coefficients of GRRC_L$_2$ are dense (Fig. 9(c)), the sample is correctly classified, as shown in Fig. 9(d).

The running time of SRC per testing sample is about 12 seconds, while GRRC_L$_1$ only needs about 1.5 seconds. However, this is still long for practical FR system. With l_2-norm regularization on the Gabor feature representation coefficients, the running time of GRRC_L$_2$ is only about 0.3 second, where 0.29 second is the running time of Gabor feature extraction. Although CRC is the fastest one, its recognition rate

is also very low, similar to that of SRC. The speedup of GRRC_L2 and GRRC_L$_1$ over SRC are 37.09 and 7.98 times, respectively.

Table 14: Recognition rates (%) and average running time (second) of GRRC and SRC on FR with disguise.

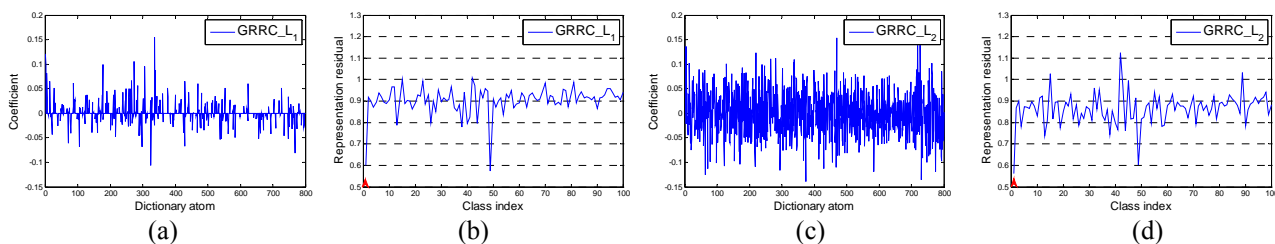|	Sunglass-S1	Scarf-S1	Sunglass-S2	Scarf-S2	Average time	Speedup
SRC	83.3	48.7	49.0	29.0	12.278	--
CRC	78.0	52.3	44.7	29.3	0.084	146.2
GRRC_L$_1$	90.7	**95.3**	50.3	**87.3**	1.539	7.98
GRRC_L$_2$	**92.3**	95	**51.7**	84.3	0.331	37.09

(a)   (b)   (c)   (d)

Figure 9: Representation coefficient and residual of a sample from class 1. (a) and (c) plot the coefficients of GRRC_L$_1$ and GRRC_L$_2$, respectively; (b) and (d) illustrate the representation residual associated to each class by GRRC_L$_1$ and GRRC_L$_2$, respectively.

It can be seen from the FR experiments with occlusion that GRRC could achieve much higher recognition accuracy than SRC and CRC. More importantly, with Gabor transformation, the occlusion dictionary could be compressed, which reduces significantly the number of unknown parameters and the computational burden. It should be noted that GRRC_L$_2$ which regularizes the coding coefficients by l_2 norm could achieve very competing performance as GRRC_L$_1$. This is because Gabor magnitude features could make original sparse representation in original image domain into a dense representation in the transformed domain.

5. Discussion of regularization on coding coefficients

In this section, we discuss the effect of feature dimension on the regularization (l_1-norm or l_2-norm) of coding coefficient. Fig. 10 plots the recognition rates of GRRC_L$_1$ and GRRC_L$_2$ versus different feature dimensionality with the same experiment setting on Mulit-PIE database in Section 4.3. The number of

dictionary atoms is 3486 (14×249). From Fig. 10, we get that when the feature dimension is too low compared to the number of dictionary atoms, GRRC_L$_1$ has better performance than GRRC_L$_2$. However, as the feature dimensionality increases, their recognition rates will become close. This result can be easily explained by the fact that when the feature dimension is much lower than the number of dictionary atoms, the dictionary is more over-complete, and thus the sparsity constraint on the representation coefficients is more reasonable. When the feature dimension is comparable to the number of dictionary atoms, especially for the problem of FR where high correlation exists in the dictionary, l_2-norm regularized GRRC has very competing performance with GRRC_L$_1$, implying that the time-consuming sparsity constraint on the coding coefficients is not necessary.
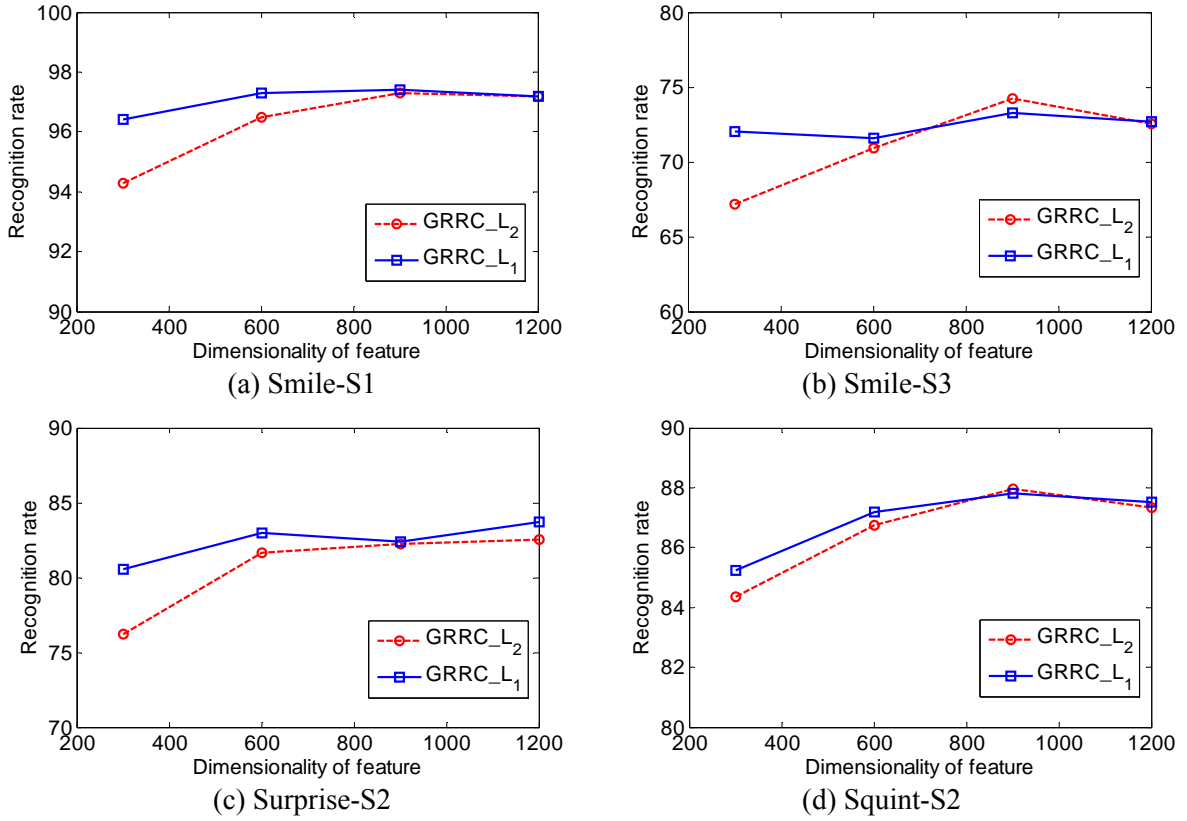

(a) Smile-S1

(b) Smile-S3

(c) Surprise-S2

(d) Squint-S2

Figure 10: Recognition rates of GRRC_L$_1$ and GRRC_L$_2$ versus feature dimensionality in FR with expression variations. (a) FR with smile in session 1 for testing. (b) FR with smile in session 3 for testing. (c) FR with surprise in session 2 for testing. (d) FR with squint in session 2 for testing.

6. Conclusion

In this paper, we proposed a Gabor-feature based robust representation and classification (GRRC) scheme for face recognition, and proposed an associated Gabor occlusion dictionary (GOD) computing algorithm to handle the occluded face images. Apart from the improved face recognition rate, one important advantage of GRRC is its compact occlusion dictionary, which has much less atoms than that of the original SRC scheme. More importantly, the coding coefficients on the learnt GOD could be regularized by l_2-norm instead of the commonly used l_1-norm. This greatly reduces the computational cost of coding. We evaluated the proposed method on different conditions, including variations of illumination, expression and pose, as well as block occlusion and disguise occlusion. The experimental results clearly demonstrated that the proposed GRRC has much better performance than SRC, leading to much higher recognition rates while spending much less computational cost. This makes it much more practical to use than SRC in real world face recognition.

References

[1] M. Kirby and L. Sirovich, "Application of the KL procedure for the characterization of human faces," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 12, no. 1, pp.103-108, 1990.
[2] M. Turk and A. Pentland, "Eigenfaces for recognition," *J. Cognitive Neuroscience*, vol. 3, no. 1, pp.71-86, 1991.
[3] P.N. Belhumeur, J.P. Hespanha, and D.J. Kriengman, "Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection," *IEEE Trans. Pattern Anal. Machine Intell.,* vol. 19, no. 7, pp. 711-720, 1997.
[4] J. Yang, J.Y. Yang, "Why can LDA be performed in PCA transformed space?" *Pattern Recognition,* vol. 36, no. 2, pp. 563-566, 2003.
[5] J.B. Tenenbaum, V. deSilva, and J.C. Langford, "A Global Geometric Framework for Nonlinear Dimensionality Reduction," *Science*, vol. 290, no. 5500, pp. 2319-2323, 2000.
[6] S.T. Roweis and L.K. Saul, "Nonlinear Dimensionality Reduction by Locally Linear Embedding," *Science*, vol. 290, no. 5500, pp. 2323-2325, 2000.
[7] X. He, S. Yan, Y. Hu, P. Niyogi, and H.J. Zhang, "Face recognition using laplacianfaces," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 3, pp. 328-340, 2005.
[8] H.-T.Chen, H.-W. Chang, and T.-L. Liu, "Local discriminant embedding and its variants," *Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition*, pp. 846-853, 2005.
[9] J. Yang, D. Zhang, J.-Y. Yang, and B. Niu, "Globally Maximizing, Locally Minimizing: Unsupervised Discriminant Projection with Applications to Face and Palm Biometrics," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 4, pp. 650-664, 2007.
[10] J. Wright, A. Yang, A. Ganesh, S. Sastry, and Y. Ma, "Robust Face Recognition via Sparse Representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp. 210-227, 2009.
[11] W. Zhao, R. Chellppa, P.J. Phillips, and A. Rosenfeld, "Face recognition: A literature survey," *ACM Computing Survey*, vol. 35, no. 4, pp. 399-458, 2003.
[12] R. Basri and D. Jacobs, "Lambertian reflectance and linear subspaces," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 25, no. 2, pp. 218-233, 2003.

[13] Y. Nesterov, A. Nemirovskii, *"Interior-point polynomial algorithms in convex programming,"* SIAM Philadelphia, PA, 1994.

[14] D. Gabor, "Theory of communication," *J. Inst. Elect. Eng.*, vol. 93, no. 26, pt. III, pp. 429-457, 1946.

[15] J.P. Jones and L.A. Palmer, "An Evaluation of the Two-Dimensional Gabor Filter Model of Simple Receptive Fields in Cat Striate Cortex", *Journal of Neurophysiology*, vol. 58, no. 6, pp. 1233-1258, 1987.

[16] M. Lades, J.C. Vorbrüggen, J. Buhmann, J. Lange, C.v.d. Malsburg, R.P. Würtz and W. Konen, "Distortion Invariant Object Recognition in the Dynamic Link Architecture", *IEEE Transactions on Computers*, vol. 42, no. 3, pp. 300-311, 1993.

[17] C. Liu and H. Wechsler, "Gabor Feature Based Classification Using the Enhanced Fisher Linear Discriminant Model for Face Recognition," *IEEE Transactions on Image Processing*, vol. 11, no. 4, pp.467-476, 2002.

[18] L. Shen and L. Bai, "A review on Gabor wavelets for face recognition," *Pattern Analysis and Application*, vol. 9, no. 10, pp. 273-292, 2006.

[19] K. Lee, J. Ho, and D. Kriegman, "Acquiring linear subspaces for face recognition under variable lighting," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 5, pp. 684-698, 2005.

[20] A.Georghiades, P. Belhumeur, and D. Kriegman, "From few to many: Illumination cone models for face recognition under variable lighting and pose," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 23, no. 6, pp. 643-660, 2001.

[21] A. Martinez and R. benavente, "The AR face database," CVC Tech. Report No. 24, 1998.

[22] P.J. Phillips, H. Wechsler, J. Huang, and P. Rauss, "The FERET database and evaluation procedure for face recognition algorithms," *Image and Vision Computing*, Vol. 16, No. 5, pp. 295-306, 1998.

[23] P.J. Phillips, H. Moon, S.A. Rizvi, and P.J. Rauss, "The FERET Evaluation Methodology for Face Recognition Algorithms," *IEEE Trans. Pattern Anal. Mach. Intell.*, Vol. 22, pp. 1090-1104, 2000.

[24] S.J. Kim, K. Koh, M. Lustig, S. Boyd, and D. Gorinevsky, "A method for large-scale l_1-regularized least squares," *IEEE Journal on Selected Topics in Signal Processing*, vol. 1, no. 4, pp. 606–617, 2007.

[25] E. Candès and J. Romberg, "L1-magic: A Collection of MATLAB Routines for Solving the Convex Optimization Programs Central to Compressive Sampling," 2006 [Online], Available: *www.acm.cal tech.edu/l1magic/.*

[26] The MOSEK Optimization Tools Version 2.5. User's Manual and Reference 2002 [Online]. Available: *www.mosek.com,MOSEK.*

[27] R. Gross, I. Matthews. J. Cohn, T. Kanade, and S. Baker, "Multi-PIE," *Image and Vision Computing*, vol. 28, pp. 807–813, 2010.

[28] S.H. Gao, I.W-H. Tsang, and L-T. Chia, "Kernel Sparse Representation for Image classification and Face recognition," *Proc. European Conf. Computer Vision*, 2010.

[29] Z. Zhou, A. Wagner, H. Mobahi. J. Wright, and Y. Ma, "Face recognition with contiguous occlusion using markov random fields," *Proc. IEEE Int'l Conf. Computer Vision*, 2009.

[30] E. Elhamifar and R. Vidal, "Robust classification using structured sparse representation," *Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition*, 2011.

[31] M. Yang, L. Zhang, J. Yang, and D. Zhang, "Robust sparse coding for face recognition," *Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition*, 2011.

[32] L. Zhang, M. Yang, and X.C. Feng, "Sparse representation or collaborative representation which helps face recognition?" *Proc. IEEE Int'l Conf. Computer Vision*, 2011.

[33] J.Z. Huang, X.L. Huang, and D. Metaxas, "Simultaneous image transformation and sparse representation recovery," *Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition*, 2008.

[34] A. Wagner, J. Wright, A. Ganesh, Z.H. Zhou, and Y. Ma, "Towards a practical face recognition system: Robust registration and illumination by sparse representation," *Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition*, 2009.

[35] S.Z. Li, "Face recognition based on nearest linear combinations," *Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition*, 1998.

[36] S.Z. Li and J. Lu, "Face recognition using nearest feature line method," *IEEE Trans. Neural Network,* vol. 10, no. 2, pp. 439-443, 1999.

[37] I. Naseem, R. Togneri, and M. Bennamoun, "Linear regression for face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 11, pp. 2106-2112, 2010.

[38] M. Yang and L. Zhang, "Gabor Feature based Sparse Representation for Face recognition with Gabor Occlusion Dictionary," *Proc. European Conf. Computer Vision*, 2010.

[39] M. Aharon, M. Elad, and A.M. Bruckstein, "The K-SVD: An Algorithm for Designing of Overcomplete Dictionaries for Sparse Representation," *IEEE Transactions on Signal Processing*, Vol. 54, no. 11, pp. 4311-4322, 2006.

[40] Q. Zhang, and B. Li, "Discriminative K-SVD for Dictionary Learning in Face Recognition," *Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition*, 2010.

[41] Z.L. Jiang, Z. Lin, L.S. Davis, "Learning A Discriminative Dictionary for Sparse Coding via Label Consistent K-SVD," *Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition*, 2011.

[42] M. Yang, L. Zhang, X.C. Feng, and D. Zhang, "Fisher Discrimination Dictionary Learning for Sparse Representation," *Proc. IEEE Int'l Conf. Computer Vision*, 2011.

[43] A.Y. Yang, A. Ganesh, Z.H. Zhou, S.S. Sastry, and Y. Ma, "A review of fast l_1-minimization algorithms for robust face recognition," arXiv:1007.3753v2, 2010.

[44] Yang and Y. Zhang, "Alternating direction algorithms for l_1-problems in compressive sensing," (preprint) arXic:0912.1185, 2009.

[45] S.F. Xie, S.G. Shan, X.L. Chen, and J. Chen, "Fusing Local Patterns of Gabor Magnitude and Phase for Face Recognition," *IEEE Trans. Image Processing,* vol. 19, no. 5, pp. 1349-1361, 2010.

[46] Q.F, Shi, and H.X. Li, "Rapid Face Recognition Using Hashing". *Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition*, 2010.

[47] H.Y. Zhou and A.H. Sadka, "Combining Perceptual Features with Diffusion Distance for Face Recognition," *IEEE Trans. System, Man, and Cybernetics-Part C: Application and Reviews*, vol. 41, no. 5, pp.577-588, 2011.

[48] S. Du and R.K. Ward, "Improved Face Representation by Nonuniform Multilevel Selection of Gabor Convolution Features," *IEEE Trans. System, Man, and Cybernetics-Part B: Cybernetics*, vol. 39, no. 6, pp.1408-1419, 2009.

[49] C.A. Perez, L.A. Cament, and L.E. Castillo, "Methodological improvement on local Gabor face recognition based on feature selection and enhanced Borda count," *Pattern Recognition*, vol. 44, pp. 951-963, 2011.

[50] S. Jahanbin, H. Choi, and A.C. Bovik, "Passive Multimodal 2-D+3-D Face Recognition Using Gabor Features and Landmark Distances," *IEEE Trans. Information Forensics and Security*, vol. 6, no. 4, pp. 1287-1304, 2011.